

Abstract

In one embodiment, the present invention extracts video regions of interest from one or more videos and generates a highly condensed visual summary of the videos. The video regions of interest are extracted based on to energy, movement, face or other object detection methods, associated data or external input, or some other feature of the video. In another embodiment, the present invention extracts regions of interest from images and generates highly condensed visual summaries of the images. The highly condensed visual summary is generated by laying out germs on a canvas and then filling the spaces between the germs. The result is a visual summary that resembles a stained glass window having cells of varying shape. The germs may be laid out by temporal order, color histogram, similarity, according to a desired pattern, size, or some other manner. The people, objects and other visual content in the germs appear larger and become easier to see. The visual summary of the present invention utilizes important regions within the key frames, leading to more condensed summaries that are well suitable for small screens.